



## A new hybrid ant colony optimization algorithm for feature selection

Md. Monirul Kabir<sup>a</sup>, Md. Shahjahan<sup>b</sup>, Kazuyuki Murase<sup>a,c,\*</sup>

<sup>a</sup> Department of System Design Engineering, University of Fukui, Fukui, Japan

<sup>b</sup> Department of Electrical and Electronic Engineering, Khulna University of Engineering and Technology (KUET), Khulna, Bangladesh

<sup>c</sup> Department of Human and Artificial Intelligence Systems and Research and Education Program for Life Science, University of Fukui, Fukui, Japan

### ARTICLE INFO

#### Keywords:

Feature selection  
Ant colony optimization  
Wrapper and filter approaches  
Hybrid search  
Neural network  
Classification accuracy

### ABSTRACT

In this paper, we propose a new hybrid ant colony optimization (ACO) algorithm for feature selection (FS), called ACOFS, using a neural network. A key aspect of this algorithm is the selection of a subset of salient features of reduced size. ACOFS uses a hybrid search technique that combines the advantages of wrapper and filter approaches. In order to facilitate such a hybrid search, we designed new sets of rules for pheromone update and heuristic information measurement. On the other hand, the ants are guided in correct directions while constructing graph (subset) paths using a bounded scheme in each and every step in the algorithm. The above combinations ultimately not only provide an effective balance between exploration and exploitation of ants in the search, but also intensify the global search capability of ACO for a high-quality solution in FS. We evaluate the performance of ACOFS on eight benchmark classification datasets and one gene expression dataset, which have dimensions varying from 9 to 2000. Extensive experiments were conducted to ascertain how ACOFS works in FS tasks. We also compared the performance of ACOFS with the results obtained from seven existing well-known FS algorithms. The comparison details show that ACOFS has a remarkable ability to generate reduced-size subsets of salient features while yielding significant classification accuracy.

© 2011 Elsevier Ltd. All rights reserved.

### 1. Introduction

Feature selection (FS) is generally used in machine learning, especially when the learning task involves high-dimensional datasets. High-dimensional datasets contain very large feature sets, which not only cause learning to be more difficult, but also degrade the generalization performance of the learned models. The purpose of FS is to simplify and enhance the quality of a dataset by selecting salient features. Ordinarily, FS deletes spurious features from the original dataset without sacrificing generalization performance. In real-world problems, FS is essential due to the existence of the following factors: (a) abundance of noise, (b) spurious information, and (c) irrelevant and redundant features in the original feature set. Accordingly, FS has become an area of active research spreading throughout many fields, including pattern recognition, data mining, image mining, and text categorization (Aghdam, Aghaee, & Basiri, 2009; Jensen, 2005).

A number of proposed approaches for FS can broadly be categorized into the following three classifications: wrapper, filter, and hybrid (Liu & Tu, 2004). In the wrapper approach, a predetermined

learning model is assumed, wherein features are selected that justify the learning performance of the particular learning model (Guyon & Elisseeff, 2003), whereas in the filter approach, statistical analysis of the feature set is required, without utilizing any learning model (Dash & Liu, 1997). Schematic diagrams of how the wrapper and filter approaches find salient features are given in Fig. 1. The hybrid approach attempts to utilize the complementary strengths of the wrapper and filter approaches (Huang, Cai, & Xu, 2007).

Subsets can be generated and the search process carried out in a number of ways. One method, called sequential forward search (SFS; Guan, Liu, & Qi, 2004; Peng, Long, & Ding, 2003), is to start the search process with an empty set and successfully add features; another option called sequential backward search (SBS; Gasca, Sanchez, & Alonso, 2006; Hsu, Huang, & Schuschel, 2002), is to start with a full set and successfully remove features. In addition, a third alternative, called bidirectional selection (Caruana & Freitag, 1994), is to start on both ends and add and remove features simultaneously. A fourth approach (Lai, Reinders, & Wessels, 2006; Stracezzi & Utgoff, 2004), is to have a search process start with a randomly selected subset using a sequential or bidirectional strategy. Yet another search strategy, called complete search (Liu & Tu, 2004), may give a best solution to an FS task due to the thoroughness of its search, but is not feasible when dealing with a large number of features. For example, assuming  $s$  to be a subset of

\* Corresponding author at: Department of Human and Artificial Intelligence Systems, University of Fukui, 3-9-1 Bunkyo, Fukui 910-8507, Japan. Tel.: +81 (0) 776 27 8774; fax: +81 (0) 776 27 8420.

E-mail address: [murase@synapse.his.u-fukui.ac.jp](mailto:murase@synapse.his.u-fukui.ac.jp) (K. Murase).

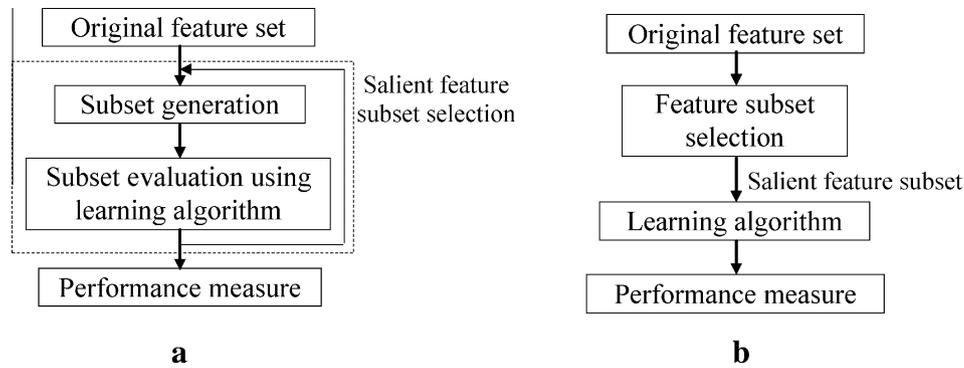


Fig. 1. Schematic diagrams of (a) wrapper approach and (b) filter approach.

selected features, and  $n$  to be the number of available features, the total computational cost of the combination is  ${}^n c_s \approx n! / (s! \cdot (n - s)!)$ . The worst case scenario occurs when the values of  $n$  and  $s$  become large. Alternatively, the sequential strategy is simple to implement and fast, but is affected by the nesting effect (Pudil, Novovicova, & Kittler, 1994), wherein once a feature is added (or, deleted) it cannot be deleted (or, added) later. In order to overcome such disadvantages of the sequential search strategy, another search strategy, called the floating search strategy (Pudil et al., 1994), has been implemented.

Most of the afore-mentioned search strategies, however, attempt to find solutions in FS that range between sub-optimal and near optimal regions, since they use local search throughout the entire process, instead of global search. On the other hand, these search algorithms utilize a partial search over the feature space, and suffer from computational complexity. Consequently, near-optimal to optimal solutions are quiet difficult to achieve using these algorithms. As a result, many research studies now focus on global search algorithms (or, “metaheuristics”) (Ke, Feng, & Ren, 2008). The significance of global search algorithms is that they can find a solution in the full search space on the basis of activities of multi-agent systems that use a global search ability utilizing local search appropriately, thus significantly increasing the ability of finding very high-quality solutions within a reasonable period of time (Dorigo & Stutzle, 2004). To achieve global search, researchers have attempted simulated annealing (Filippone, Masulli, & Rovetta, 2006), genetic algorithm (Yang & Honavar, 1998), ant colony optimization (Aghdam et al., 2009; Ani, 2005), and particle swarm optimization (Wang, Yang, Teng, Xia, & Jensen, 2007) algorithms in solving FS tasks.

In this paper, we propose a new hybrid ant colony optimization (ACO)-based FS algorithm (ACOFS) that utilizes a hybrid search strategy in the feature space. The idea incorporated in this algorithm was originally introduced in our earlier work (Kabir, Shahjahan, & Murase, 2009). The main focus of this algorithm is to generate subsets of salient features of reduced size. The proposed method utilizes a hybrid search technique that combines the wrapper and filter approaches. It has often been found that hybrid techniques are capable of finding a good solution, even when a single technique is often trapped with an incomplete solution. In this regard, ACOFS modifies the standard pheromone update and heuristic information measurement rules based on the above two approaches. The reason for the novelty and distinctness of the proposed ACOFS versus previous algorithms (Aghdam et al., 2009; Ani, 2005; Kanan, Faez, & Taheri, 2007; Ke et al., 2008; Khushaba, Alsukker, Ani, & Jumaily, 2008; Robbins, Zhang, & Bertrand, 2008; Sivagaminathan & Ramakrishnan, 2007) lie in the following two aspects.

First, ACOFS emphasizes not only the selection of a number of salient features, but also the attainment of a reduced number of them. ACOFS selects salient features of a reduced number using a subset size determination scheme. Such a scheme works upon a bounded region and provides sizes of constructed subsets that are smaller in number. Thus, following this scheme, an ant attempts to traverse the node (or, feature) space to construct a path (or, subset). This approach is quite different from those of the existing schemes (Aghdam et al., 2009; Kanan et al., 2007; Ke et al., 2008), where the ants are guided by using the SFS strategy in selecting features during the feature subset construction (SC). However, a problem is that, SFS requires an appropriate stopping criterion to stop the SC. Otherwise, a number of irrelevant features may be included in the constructed subsets, and the solutions may not be effective. To solve this problem, some algorithms (Ani, 2005; Khushaba et al., 2008; Robbins et al., 2008; Sivagaminathan & Ramakrishnan, 2007) define the size of a constructed subset by a fixed number for each iteration for all ants, which is incremented at a fixed rate for following iterations. This technique could be inefficient if the fixed number becomes too large or too small. Therefore, deciding the subset size within a reduced area may be a good step for constructing the subset while the ants traverse through the feature space.

Second, ACOFS utilizes a hybrid search technique for selecting salient features that combines the advantages of the wrapper and filter approaches. An alternative name for such a search technique is “ACO search”. This technique is designed with two sets of new rules for pheromone update and heuristic information measurement. The idea of these rules is based mainly on the random and probabilistic behaviors of ants while selecting features during SC. The aim is to provide the correct information to the features and to maintain an effective balance between exploitation and exploration of ants during SC. Thus, ACOFS achieves a strong search capability that helps to select a smaller number of the most salient features among a feature set. In contrast, the existing algorithms (Aghdam et al., 2009; Ani, 2005; Kanan et al., 2007; Ke et al., 2008; Khushaba et al., 2008; Robbins et al., 2008; Sivagaminathan & Ramakrishnan, 2007) try to design rules without distinguishing between the random and probabilistic behaviors of ants during the construction of a subset. Consequently, ants may be deprived of the opportunity of utilizing enough previous experience or investigating more salient features during SC in their solutions.

The rest of this paper is organized as follows. Section 2 describes some related works concerning FS. A detailed description of our proposed ACOFS, including the computational complexity of different stages, is described in detail in Section 3. Section 4 presents the results of our experimental studies, including the methodology, results, and a comparison with other existing FS

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات